**MEMORANDUM | September 1, 2015**

**TO** Craig O'Connor, NOAA

**FROM** Eric English

**SUBJECT** G3 – Local Shoreline and Boating Valuation Weights

## 1. THE LOCAL VALUATION SURVEY

The local valuation survey collected information about recreational shoreline and boating trips to coastal locations in the six states of Texas, Louisiana, Mississippi, Alabama, Florida, and Georgia. Respondents to the survey were drawn from a coverage area that included the same six states, but excluded parts of Georgia and Texas that were covered in the National Valuation Survey. A given respondent was asked either about boating trips or about shoreline trips, where shoreline trips involved any coastal recreation activities other than boating, such as sunbathing, swimming, or fishing.

The survey was conducted in 12 successive releases, or waves. Each wave consisted of an initial contact by mail and a follow-up telephone interview. The mail survey was only a single page and included a question about the household's participation in coastal recreation over the previous year. All households that returned the mail survey and said they had engaged in coastal recreation during the previous year were included in the telephone portion of the survey. A fraction of households that had not participated in coastal recreation were subsampled and were also included in the telephone survey.

The telephone survey asked detailed questions about the respondent's recreation trips in the previous two to four months. Specifically, the first-wave interviews were conducted during June and July of 2012 and asked respondents about their recreation from April 2012 up to the time of the survey. The second wave was conducted during July and August 2012 and asked respondents about their recreation from May 2012 up to the time of the survey, and so on.

Two valuation data sets were obtained using the results of the local survey. The boating valuation data includes information provided by respondents about their boating trips, and the shoreline valuation data includes information provided by respondents about their shoreline trips. Respondents who did not take boating or shoreline trips were included as nonparticipants in both data sets. Given the development of appropriate sampling weights, each data set is representative of the entire population of individuals age 18 or older living in the specified six-state region.

This memorandum describes the development of weights for the shoreline and boating data sets. The Technical Memo G1 – Local Valuation Survey describes details of the local survey.

## 2. FEATURES OF THE SURVEY INCORPORATED INTO THE WEIGHTS

The following features of the local valuation survey and sample design were important in developing the weights.

- Home addresses were sampled from two sample frames: a list of all household addresses in the coverage area compiled by the United States Postal Service, and a list of addresses for all boats registered in the coverage area. Lists of registered boats were provided by the six states. The first frame is referred to as the Address-Based Sampling (ABS) frame, and the second frame is referred to as the boat frame. While a household can have more than one boat, duplicate addresses were removed from the boat frame prior to sampling.
- From each frame, a sample was drawn for the mail survey consisting of a random selection of home addresses stratified by state and county. The size of the sample in each county was proportionate to the square root of the total population of each county, multiplied by a factor specific to each of eight regions. The factors for the eight regions were 0.5 for Georgia, Texas, and the Atlantic side of the Florida Peninsula; 0.75 for Mississippi; and one for the remaining areas, including the Gulf side of the Florida Peninsula, the Florida Panhandle, Louisiana, and Alabama. The total sample size for the ABS frame was 325,285 and the total size of the sample from the boat frame was 67,072.
- Respondents to the mail portion of the survey who indicated they had participated in coastal recreation in the previous year were selected into the telephone survey. Respondents to the mail survey who had not participated in coastal recreation in the previous year were subsampled for the telephone survey at a rate of 0.13 if they were selected from the ABS frame and at a rate of 0.35 if they were selected from the boat frame.
- Respondents to the telephone survey were selected for one of three paths in the telephone survey based on their participation in coastal recreation activities during the period addressed in the interview. Those who had participated in shoreline activities only were selected for the path involving shoreline activities and demographic questions. Those who had participated in boating activities only were selected for the path involving boating activities and demographic questions. Those who had not participated in coastal recreation were selected for the path involving only demographic questions. Those who had participated in both shoreline and boating activities were selected for the boating path with a 0.75 probability and for the shoreline path with a 0.25 probability.
- The purpose of the survey was to compile two data sets, one for boating activities and one for shoreline activities. Each of the data sets represents activities over a 12-month period for all residents at least 18 years old living in the six-state coverage area. Because respondents in each of the 12 monthly survey waves reported activities for a two-to-four-month period, the combination of data from the 12 waves resulted in overlap in the reporting periods across waves. To construct the final weighted data, activities were represented each month by a separate sample, consisting of all respondents who

reported about their activities for that month. The sample of respondents reporting in a given month was weighted up to represent the full population using raking procedures, as described below.

## 3. DEVELOPMENT OF THE WEIGHTS

The steps below describe the development of weights for the local shoreline and boating valuation survey.

- **Calculate base weights using sample selection probabilities**. The base weight for each record involves the inverse of the selection probabilities for the sampled addresses. However, adjustments were required to account for the use of a dual ABS and boat frame. First, a selection probability was calculated for addresses sampled from each frame. The selection probability $\pi_{hk}$ for address k sampled from frame h in a given county was equal to the number of addresses sampled in the county divided by the number of addresses in the frame for that county. Second, addresses sampled from the boat frame that were also sampled in the ABS frame were dropped from the boat-frame sample. Base weights were then calculated differently for each of three groups. The first group included addresses sampled from the ABS frame that did not match any addresses in the boat frame. Addresses in this group were assigned a base weight of $d_{1k} = 1/\pi_{1k}$, where h = 1 refers to cases selected from the ABS frame. The second group was addresses sampled from the boat frame, which were assigned a base weight $d_{2k} = 1/\pi_{2k}$, where h = 2 represents cases selected from the boat frame. The third group was addresses selected from the ABS frame that matched an address in the boat frame. These cases were assigned a weight of $d_{1k} = 1/( \pi_{1k} + \pi_{2k})$.

- **Apply a nonresponse adjustment for the mail survey**. Mail survey records fall into one of four categories: 1) those who completed the mail survey (C); 2) those who returned the mail survey but indicated they did not want to participate in the survey (R); 3) those who were ineligible for the mail survey, consisting of invalid or non-residential addresses (I); and 4) those who did not return the mail survey and therefore have unknown status (U).

  The categories C, R, I, and U were defined separately for each of 24 weighting classes, which are the cross-classification of the six states, the two frames, and a binary indicator for coastal counties. The nonresponse adjustment consists of two factors, calculated separately for each of the weighting classes. Let S represent the sum of the base weights for records in a given response category and a given weighting class (e.g. for each weighting cell L, $S_{CL} = \Sigma_{k \epsilon C \epsilon L} d_k$). Ignoring the subscript L, the first adjustment factor is

$$A = \frac{S_C + S_R + S_I + S_U}{S_C + S_R + S_I}.$$

  This factor was multiplied by the base weight for each record in C + R + I, while records in U received a weight of zero. This factor distributed the weight

of all records in U, with unknown status, to records in the three remaining categories with known status.

Now let S represent the sum of weights in the same categories as before, but the weights now include the effect of the first adjustment factor. The second factor is

$$B = \frac{S_C + S_R}{S_C}.$$

This factor was applied to each record in C, while records in R and I received a weight of zero. This adjustment eliminated refusals and ineligibles while weighting up completed surveys to represent C + R, the full eligible population.

- **Adjust for the subsampling of recreation nonparticipants**. Respondents to the mail survey whose household had not participated in coastal recreation during the previous year were subsampled for the telephone survey. For cases selected from the ABS frame, subsampling of mail nonparticipants occurred at a rate of 0.13, and the selected cases were therefore weighted up by a factor of 1/0.13. For the boat frame, the subsampling rate was 0.35, and the weight for selected cases was 1/0.35.
- **Post-stratify to household counts**. At this stage all weights for recreation participants and selected nonparticipants were rescaled so that the sum of the weights in each of the six states matched the total number of households in each state as reported in the 2010 Census. The adjustment factor for a record in state m is

$$H_M = \frac{T_M}{\sum_{k \in M} w_k}.$$

$T_M$ is the total number of households in state M, or in the case of Georgia and Texas, the total number of households in the relevant counties. The weights $w_k$ incorporate all adjustments described previously. This is the final adjustment at the household level.

- **Adjust for sub-sampling of adults within a household.** At the start of a telephone interview, a respondent was selected at random from among all members of the household who were 18 or older. Each weight was therefore multiplied by an adjustment factor equal to the number of adults in the household. The weights for those who did not respond to the telephone interview were set to zero.
- **Impute for missing values to be used in raking.** In preparation for raking the weights, a hot-decking procedure filled in any missing values for demographic variables used in the raking. The variables were age (18 to 24, 25 to 34, 35 to 44, 45 to 64, and 65 or older), education (high school or less, some college, bachelor's or higher), race (Hispanic of any race, Black non-Hispanic, other non-Hispanic), and sex. The hot-decking procedure involved dividing the sample into groups, or "cells". Within each cell, a missing value for a given record was filled in using a randomly selected "donor" record from within the

same cell. The variables age, sex, and race were filled in first, using 12 cells based on the six states crossed with the two frames. Education was then imputed using 30 cells, formed by crossing the five age levels with the six states.

- **Adjust for the selection of the boating or shoreline interview path, while creating separate data sets for the boating and shoreline activities.** Respondents who had participated in both shoreline and boating activities during the period addressed in the telephone interview were randomly assigned to one of two survey paths. The boating path involved questions about a respondent's boating activities, and the shoreline path involved questions about a respondent's shoreline activities. Since the probability of selection for the boating path was 0.75, respondents selected for the boating path received an additional weight of 4/3. Respondents selected for the shoreline path were assigned an additional weight of four.

  At this point two separate data sets were created. One data set consists of all respondents except those who had participated in both activities and were randomly selected for the shoreline path. This data set represents boating activities for all residents of the survey coverage area. Those who participated only in shoreline activities are included in the boating data set and considered nonparticipants. The other data set consists of all respondents except those who had participated in both activities and were randomly selected for the boating path. This data set represents shoreline activities for all residents of the survey coverage area. In the shoreline data set, the set of nonparticipants includes those who participated only in boating.

- **Rake each of the two data sets to control totals from the 2010 Census.** For the boating and shoreline data sets, respectively, respondents across all survey waves were combined and raked to control totals using four raking dimensions. Each dimension involved dividing the sample into cells that could be matched to control totals from the 2010 Census. The four dimensions were 1) education crossed with race, resulting in nine cells; 2) age by education, with 15 cells; 3) sex by education, with six cells; and 4) state by race, with 18 cells.

  The raking procedure began by calculating control totals for each cell in each dimension. These are just the total number of people in the U.S. population over 18 falling within each cell, according to the 2010 Census. The next step was to sum the sample weights in each cell of the first dimension. Each sample weight in each cell was then multiplied by the ratio of the control total from the Census to the sum of the weights in the cell. The procedure was repeated for the remaining three dimensions, each time beginning with the adjusted weights from the previous step. Cycling through the four dimensions one time represents one iteration. Iterations were repeated until changes in the weights at each iteration fell below a selected convergence criterion.

- **Trim and re-rake.** The weights were grouped into cells, and large weights within each cell were trimmed. For this step there were 48 cells, formed by the cross-classification of the six states, the two frames, a binary indicator for coastal counties, and the respondent's participation status during the previous year as determined in the mail survey. The determination of the "large weights" threshold was based on procedures developed in the sampling literature. Trimmed weights were reduced to the size of the largest untrimmed weight. After trimming, the weights were re-raked to again match the control totals. At this stage the weights represent the final "annual" weights, which are used for sensitivity analysis. An adjustment to weight up each respondent's reporting period of a few months to a full year is required when using the annual weights.

- **Rake to control totals by month to create monthly weights.** The final weights used in the local shoreline and boating valuation models are monthly weights, which are used to generate a representative sample of boating or shoreline activities for each month of the year. The creation of monthly weights was performed separately for the boating and shoreline data sets.

  First, the set of respondents reporting their activity in a given month was identified. Any respondent reporting about their activities for a period of at least 10 days in a given month was considered to be a respondent for that month. For example, a respondent contacted on July 15 might have been asked to report activities since May 1. This respondent would have reported activities for all of May and June, and for 15 days in July. This respondent would be part of the set of respondents used to develop weights for each of these three months. A respondent reporting for the period May 1 to July 5 would be in the set of respondents for May and June only.

  The set of respondents for a given month was then raked to control totals for the full population. The inputs to the raking process were the annual weights calculated above. The dimensions for the monthly raking were 1) the six states; 2) the five age categories; 3) the 10 categories created by crossing age and sex; and 4) the three race categories. The raking process not only adjusted the sample of respondents in each month to match the proportions in the raking dimensions. The raking also rescaled the weights to represent the full population, given that the set of respondents in each month was only a subset of all respondents included in the annual weights.

  For any respondent reporting for only a portion of days in a given month, a factor was created to expand the reporting period in that month to a full month. The factor was the number of days in the month divided by the number of days in the respondent's reporting period for that month.

- **Generate replicate weights for variance estimation.** A set of 120 replicate weights was created for use in variance estimation. Starting with the original single data set and the original base weights, all sampled addresses were sorted by frame, state, and county. The first two of the sorted records were considered

a pair in group 1, the next two records were a pair in group 2, and so on, for the first 120 pairs. After that the numbering began again, so that the 121st pair was also in group 1. This process continued until all record pairs were placed in one of 120 groups.

One set of replicate weights was created for each of the 120 groups. To generate the first set of replicate weights, one record in each pair of records in group 1 was eliminated. The weight of the other record in each pair was doubled. The weights for all records outside of group 1 were left unchanged. The same process was applied to each of the remaining groups of records to form 120 sets of replicate base weights. Each set of replicate base weights was then adjusted using the full set of steps described above, from the nonresponse adjustment to the raking and trimming procedures. The result was 120 sets of final replicate weights.